

# 人工知能的手法を採用したリアルタイムモニタリングシステムの提案と検証

## FPGA-based video monitoring using an AI-based approach

李成哲\*<sup>1</sup>  
Li Chengzhe

赖荣亿\*<sup>1</sup>  
Lai Yoong Yee

山口佳樹\*<sup>1,2</sup>  
Yoshiki Yamaguchi

川本 雅之\*<sup>1</sup>  
Masayuki Kawamoto

\*<sup>1</sup>筑波大学  
University of Tsukuba

\*<sup>2</sup>産業技術総合研究所人工知能研究センター  
Artificial Intelligence Research Center, AIST

High-performance and energy-efficient stream processing is focused for accelerating the revolution of Internet of Things (IoT) and Machine 2 Machine (M2M). Looking at video monitoring in the IoT/M2M world, sophisticated processing such as video semantic analysis is required strongly though most current systems can only compress recorded video in real-time. It will enable organized highly efficient operation of IoT/M2M system because each node works as strong decentralized processing system and avoids unnecessary network load. This paper proposes an FPGA-based system that achieves sufficient computing performance and sophisticated computation. The proposed prototype system is able to recognize the action of six volleyball players in real-time. An AI approach has been used for player recognition and it contributes to reduce the computational effort and the size of a semantic image database. Through the evaluation of volleyball monitoring on a real system, the result showed that the combination of AI and FPGA-based acceleration can achieve high-performance and high-accuracy computing.

### 1. はじめに

半導体製造技術の向上により、撮像素子の性能向上および小型化・軽量化が実現され、携帯機器による動画情報処理が一般に受け入れられつつある。つまり、IoT/M2M 社会において、録画・圧縮・保存・再生といった従来型の動画処理だけではなく、動画の持つ意味をリアルタイムで解析することが求められる。より具体的には、車載システム [Ishiduka2004]、監視システム [Altera2007]、スポーツ分野 [Takahashi2014]、データベース処理 [Putnam2014] などである。

意味解析では、意味というものを何らかの演算により求め、システム内で定量的に評価する必要がある。深層学習による画像解析 [Le2012] などは大変良い例である。一方、実時間動画処理にこの意味解析処理を導入し、かつ、携帯可能なシステムで実現するのは大変難しい。これは、演算性能もだが、OS を伴ったいわゆる IT 機器では録画装置より連続して送られてくる動画(ストリームデータ)を効率良く処理するのに向いていないからである。小規模なシステムで意味解析を実現するためには、演算の全てがパイプライン化され、ストリーム処理に特化した加速装置について考慮する必要がある。

そこで本研究では、画像処理については従来手法を、意味解析には人工知能的手法を用いてこの問題解決に取り組む。全ての演算は基本的に、Field Programmable Gate Array (FPGA) と呼ばれる電氣的に回路情報を書き換え可能なハードウェアを用いて実現する。

次に、試作システムの検証用アプリケーションとして、バレーボールにおける選手の実時間認識を選択した。提案システムでは、6人の選手の動きについて特徴抽出・認識処理・追尾処理を録画速度(30fps)以上の速度で実現する。

最後に、本提案はセンサー端末を含む携帯用情報端末などにおいて、録画情報から得られる重要な情報や動画のたじゆ情報のみ送信することを可能にする。つまり、通信の高速化やネットワーク帯域およびストレージ容量の節約などにも応用できる。

### 2. システムの概要

本研究では、10cm<sup>3</sup> 規模が携帯可能なサイズの限界と考えている。そこで、カメラ・演算部・電源等を含めてこのサイズに含まれるようにシステム設計を行った。図 1 に、提案するシステムの構成要素と今回対象としたアプリケーション(バレーボール)を示す。

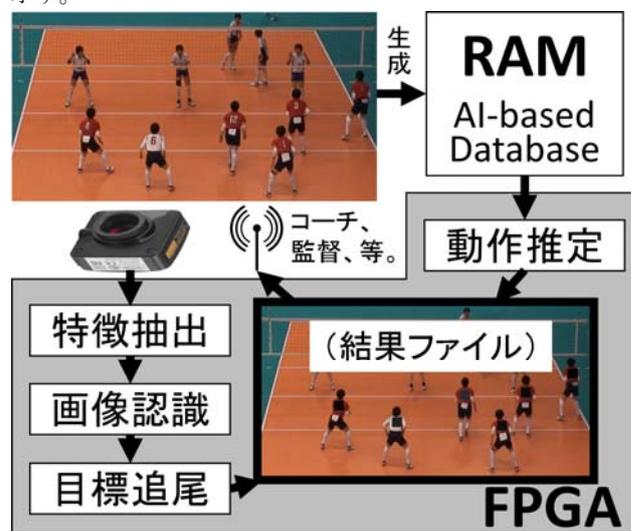


図 1 提案システムの概要と対象としたアプリケーション

図 1 のシステムにおいて、まずカメラから映像を直接 FPGA に入力する。FPGA は、3つの演算(特徴抽出、画像認識、目標追尾)について 33ms 以内で処理するように設計されている。また、これとは別に画像分類用のメモリも保持しており、これによってより高度な分類や認識処理を可能としている。演算結果(図1では背番号部分がハイライトされている)は、随時出力される。バレーボールの例であれば、選手の移動軌跡や動作などが時間情報とともに記録することが可能となり、電子スコアラーとしての利用が期待できる。

### 3. 実システム検証環境

図 2 に、本研究における試作システムおよび検証環境について示す。検証において、撮影した試合を一方のディスプレイで再生する。再生された動画は FPGA に接続されたカメラで改めて撮影され、FPGA ボードで必要な処理を適応後、その演算出力をもう一面のディスプレイに表示する。

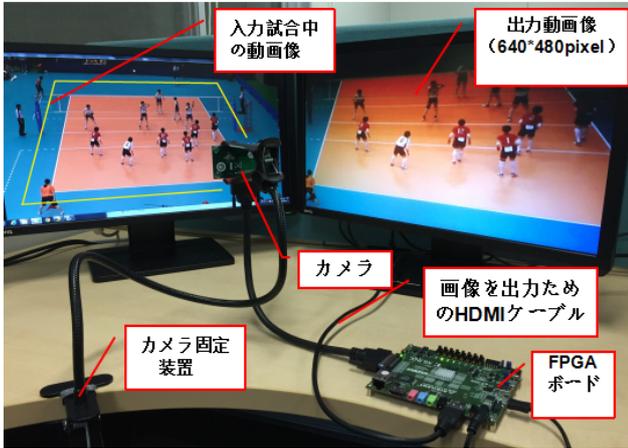


図 2 試作システムおよび検証環境

検証用環境の整備により、問題の再現や特徴的な出力の整理などが可能となった。一方、画面のちらつきやリフレッシュレートの影響など、実世界では発生しないノイズにより精度が低下するという問題も発生した。これを削減する為、1ms の高速応答、144Hz 駆動、フリッカーフリーバックライトを採用している液晶ディスプレイ(BenQ 社製 XL2411Z)を使用している。

### 4. 実験結果

#### 4.1 ハードウェア使用量

表 1 にハードウェア使用量を示す。FPGA ボードは Digilent 社製 Atlys Spartan-6 FPGA Trainer Board を使用した。このときの回路記述は Verilog-HDL を使用し、CAD ツールには Xilinx 社の ISE Design Suite 14.6 を用いている。

表 1 回路使用量(Spartan-6 LX45FPGA における使用率)

	回路使用量(LUTs)	メモリブロック数(BRAM)
選手認識	6,522(23%)	3(2.6%)
選手追跡	8,049(29%)	7(6.0%)

#### 4.2 選手情報の更新

提案システムは単眼であるため、選手の重なりなどの状態に対応するのが極めて難しい。そこで、図 3 に示す様に選手の推定処理も合わせて行い、選手位置情報等の更新も行っている。



図 3 追跡中の重なりにおける情報更新(連続 5 フレーム)

### 4.3 リアルタイム処理における認識率

バレーボールの 1 ラリー(約 7.7 秒、フレーム枚数 184 枚)における選手 6 名の認識率は実機において約 66% であった。

性能低下の理由には、画面のちらつきや撮影画面が近いことによる対象物体の歪み(急激な形状変化)等も含まれていた。これらのノイズ低減処理および動画情報という特徴を活かした手法(前後フレームにおける認識結果の再利用)により、92%程度まで精度を改善できると著者らは試算している。現在はこの試算に基づき、FPGA における実システムの改良を進めている。

### 5. おわりに

小〜中規模に属する廉価な FPGA ボードを利用し、6 名の選手の認識・追跡処理を実時間で処理できることを確認した。現在の認識率は 65%程度と高くないが、90%以上まで改善できることは確認されており、実利用にも耐えるシステムであると考えている。

今後は、精度の改善だけでなく、部分空間法によるデータベース認識に加え、使用していない回路部分に 2 値化ニューラルネットワーク[Courbariaux2016] [Smaragdis2015] [Zhang2015]などを導入することを計画している。ニューラルネットワークによる空間並列処理と FPGA は親和性が高いため、コンパクトでありながら効率の高い回路が実現できると考えている。

### 6. 参考文献

- [Ishiduka2004] 石塚裕, 平井有三, "Opponent-Color フィルタを用いた道路交通標識認識システム", 信学技報, パターン認識・メディア理解 103(737), pp.13-18,2004.
- [Altera2007] Altera Corporation, "FPGA を使用したビデオ監視の実装", Altera Corporation White Paper, 2007.
- [Takahashi2014] 高橋正樹, 中村俊之, 三科智之, "機械学習を利用した複数視点映像からのサッカーボール追跡", 映像情報メディア学会技術報告 38(51), pp.5-8, 2014.
- [Putnam2014] A. Putnam, A. Caulfield and E. Chung, "Large-Scale Reconfigurable Computing in a Microsoft Datacenter", Hot Chips 26, 2014.
- [Le2012] Q. Le, M. Ranzato, R. Monga, M. Devin, K. Chen, G. Corrado, J. Dean, A.Y. Ng, "Building High-level Features Using Large Scale Unsupervised Learning", International Conference in Machine Learning, 2012.
- [Courbariaux2016] M. Courbariaux, I. Hubara, D. Soudry, R. El-Yaniv and Y. Bengio, "Binarized Neural Networks: Training Deep Neural Networks with Weights and Activations Constrained to +1 or -1", Computing Research Repository, arXiv:1602.02830v3, pp.1-7, March 2016.
- [Smaragdis2015] M. Kim and P. Smaragdis, "Bitwise Neural Networks", In Proc. Int'l Conf. on Machine Learning Workshop on Resource-Efficient Machine Learning, pp.1-5, July 2015
- [Zhang2015] J. Zhang, M. Utiyama, E. Sumita, G. Neubig and S. Nakamura, "A Binarized Neural Network Joint Model for Machine Translation", In Proc. 2015 Conference on Empirical Methods in Natural Language Processing, pp. 2094-2099, September 2015.